



# **Development of a High-Resolution Binational Vegetation Map of the Santa Cruz River Riparian Corridor and Surrounding Watershed, Southern Arizona and Northern Sonora, Mexico**

By Cynthia S.A. Wallace, Miguel L. Villarreal, and Laura M. Norman

Open-File Report 2011–1143

U.S. Department of the Interior  
U.S. Geological Survey

**U.S. Department of the Interior**  
KEN SALAZAR, Secretary

**U.S. Geological Survey**  
Marcia K. McNutt, Director

U.S. Geological Survey, Reston, Virginia: 2011

For more information on the USGS—the Federal source for science about the Earth, its natural and living resources, natural hazards, and the environment—visit <http://www.usgs.gov> or call 1-888-ASK-USGS

For an overview of USGS information products, including maps, imagery, and publications, visit <http://www.usgs.gov/pubprod>

Suggested citation:

Wallace, C.S.A., Villarreal, M.L., and Norman, L.M., 2011, Development of a high-resolution binational vegetation map of the Santa Cruz River riparian corridor and surrounding watershed, southern Arizona and northern Sonora, Mexico: U.S. Geological Survey Open-File Report 2011-1143, 22 p. [<http://pubs.usgs.gov/of/2011/1143/>].

Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Although this report is in the public domain, permission must be secured from the individual copyright owners to reproduce any copyrighted material contained within this report.

## Contents

Conversion Factors .....	iv
Abbreviations.....	iv
Abstract .....	1
Introduction.....	1
Objective .....	4
Model Inputs.....	5
Independent Data .....	5
Training Data .....	10
Methods.....	11
Independent Data Preparation.....	11
Training Data Preparation.....	11
Vegetation Classification .....	13
Accuracy Assessment .....	14
Results.....	15
Conclusions.....	19
References .....	20

## Figures

1. Southwest Regional Gap Analysis Project (SWReGAP) map for the analysis area surrounding the Santa Cruz Watershed.....	3
2. Mapped riparian vegetation classes .....	4
3. Study area.....	5
4. Spring Landsat.....	7
5. Summer Landsat.....	8
6. Fall Landsat .....	8
7. Digital Elevation Model (DEM) of the study area and its derivatives .....	9
8. Landforms of the Santa Cruz Watershed.....	10
9. Flow Accumulation and derivatives.....	10
10. Southwest Regional Gap Analysis Project (SWReGAP) core areas .....	12
11. Training points .....	13
12. Santa Cruz Watershed map of the riparian corridor and surrounding watershed .....	16
13. Vegetation classifications in three maps of the riparian corridor.....	16
14. Selected Terrestrial Ecological Systems units in the Santa Catalina Mountains .....	17

## Tables

1. Error matrix for CART modeled vegetation classes.....	18
2. Classification cross-walk.....	19

## Map Raster Metadata and Data

*Metadata* (TXT file, 25 kB)

*Raster data* (ZIP file, 1.4 MB compressed; 19 MB expanded)

## Conversion Factors

Multiply	By	To obtain
kilometer (km)	0.6214	mile (mi)
meter (m)	3.281	foot (ft)

## Abbreviations

<b>AOI</b>	area of interest
<b>B</b>	blue
<b>CART</b>	Classification And Regression Tree
<b>DEM</b>	Digital Elevation Model
<b>DOQQ</b>	Digital Orthophoto Quarter Quadrangle
<b>EPM</b>	Ecosystem Portfolio Model
<b>EROS</b>	Earth Resources Observation and Science
<b>G</b>	green
<b>GAP</b>	Gap Analysis Program
<b>GIS</b>	geographic information system
<b>GLOVIS</b>	Global Visualization Viewer
<b>MIR</b>	middle-infrared
<b>NAIP</b>	National Agricultural Imagery Program
<b>NDVI</b>	Normalized Difference Vegetation Index
<b>NED</b>	National Elevation Dataset
<b>NIR</b>	near-infrared
<b>NLCD</b>	National Land Cover Database
<b>NM-GAP</b>	New Mexico GAP
<b>PCA</b>	Principal Components Analysis
<b>R</b>	red
<b>SCWEPM</b>	Santa Cruz Watershed Ecosystem Portfolio Model
<b>SWIR</b>	shortwave infrared
<b>SWReGAP</b>	Southwest Regional Gap Analysis Project
<b>TES</b>	Terrestrial Ecological Systems
<b>TM</b>	Thematic Mapper
<b>TOA</b>	top of atmosphere
<b>USGS</b>	U.S. Geological Survey

# Development of a High-Resolution Binational Vegetation Map of the Santa Cruz River Riparian Corridor and Surrounding Watershed, Southern Arizona and Northern Sonora Mexico

By Cynthia S.A. Wallace<sup>1</sup>, Miguel L. Villarreal<sup>2</sup>, and Laura M. Norman<sup>1</sup>

## Abstract

This report summarizes the development of a binational vegetation map developed for the Santa Cruz Watershed, which straddles the southern border of Arizona and the northern border of Sonora, Mexico. The map was created as an environmental input to the Santa Cruz Watershed Ecosystem Portfolio Model (SCWEPM) that is being created by the U.S. Geological Survey for the watershed. The SCWEPM is a map-based multicriteria evaluation tool that allows stakeholders to explore tradeoffs between valued ecosystem services at multiple scales within a participatory decision-making process. Maps related to vegetation type and are needed for use in modeling wildlife habitat and other ecosystem services. Although detailed vegetation maps existed for the U.S. side of the border, there was a lack of consistent data for the Santa Cruz Watershed in Mexico. We produced a binational vegetation classification of the Santa Cruz River riparian habitat and watershed vegetation based on NatureServe Terrestrial Ecological Systems (TES) units using Classification And Regression Tree (CART) modeling. Environmental layers used as predictor data were derived from a seasonal set of Landsat Thematic Mapper (TM) images (spring, summer, and fall) and from a 30-meter digital-elevation-model (DEM) grid. Because both sources of environmental data are seamless across the international border, they are particularly suited to this binational modeling effort. Training data were compiled from existing field data for the riparian corridor and data collected by the NM-GAP (New Mexico Gap Analysis Project) team for the original Southwest Regional Gap Analysis Project (SWReGAP) modeling effort. Additional training data were collected from core areas of the SWReGAP classification itself, allowing the extrapolation of the SWReGAP mapping into the Mexican portion of the watershed without collecting additional training data.

---

<sup>1</sup>U.S. Geological Survey.

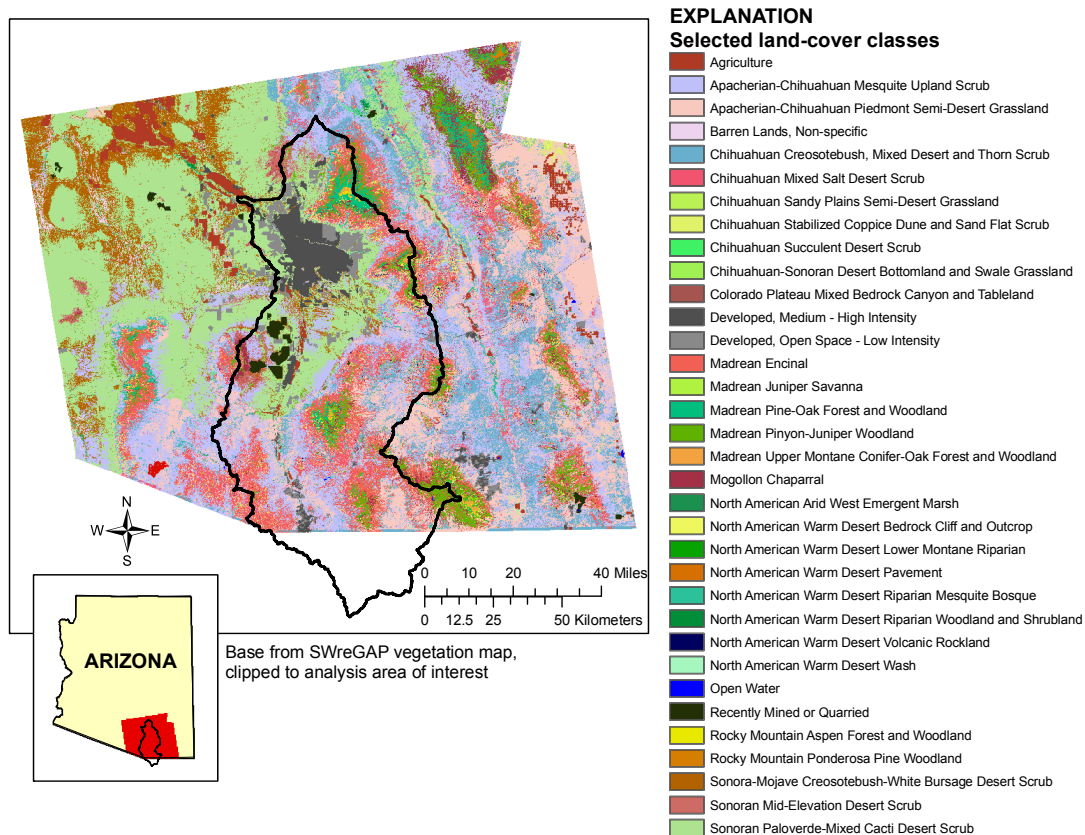
<sup>2</sup>University of Arizona.

## Introduction

A modeling tool created by the U.S. Geological Survey (USGS), the Ecosystem Portfolio Model (EPM; Labiosa and others, 2009), is being developed for the Santa Cruz Watershed (Norman and others, 2010), which straddles the southern border of Arizona and the northern border of Sonora, Mexico. The EPM is a map-based multicriteria evaluation tool that stakeholders can use to explore tradeoffs between valued ecosystem services at multiple scales within a participatory decision-making process. Maps related to vegetation type and condition that span the international border are needed for input to the Santa Cruz Watershed Ecosystem Portfolio Model (SCWEPM) for use in modeling wildlife habitat and other ecosystem services. Although detailed vegetation maps exist for the U.S. side of the border, there is a lack of consistent data for the Santa Cruz Watershed in Mexico. This report summarizes the development of a binational vegetation map.

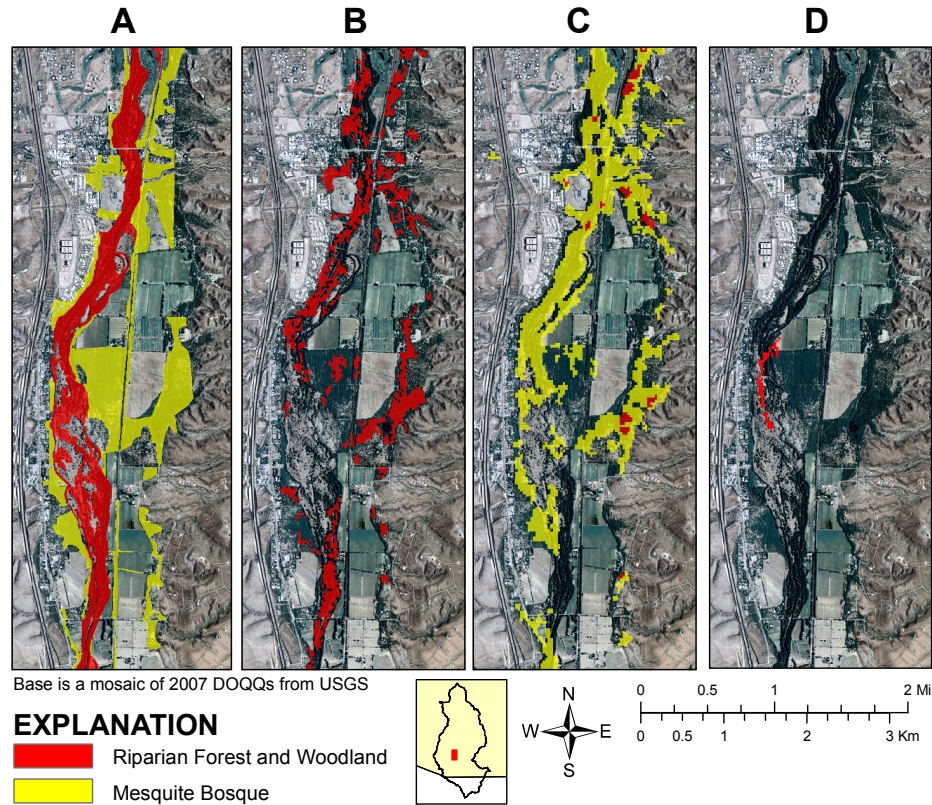
The Southwest Regional Gap Analysis Project (SWReGAP) vegetation map for the U.S. portion of the Santa Cruz Watershed was used as a starting point for this mapping effort (fig. 1; Lowry and others, 2006). The SWReGAP is part of the USGS Gap Analysis Program (GAP) that couples animal habitat models with land stewardship to identify conservation opportunities in the form of gaps in the spatial extent of important habitat for various common species (Lowry and others, 2006). The SWReGAP modeled vegetation, which was used as input to the habitat models, by using a Classification And Regression Tree (CART) algorithm (Hansen and others, 1996; Lawrence and Wright, 2001), with several environmental data layers (including seasonal Landsat images and topography) used as predictors and field data used to train the classifier. For the SWReGAP mapping effort, NatureServe (<http://www.natureserve.org/>) developed the Terrestrial Ecological Systems (TES) Classification (Grossman and others, 1998; Lowry and others, 2006) that serves as a regionally consistent mesoscale land-cover classification between the more general formation (physiognomically defined) and the more detailed alliance (floristically defined) mapping levels (Grossman and others, 1998). This mesoscale classification is sufficiently general to be applicable on regional scales while providing enough detail to be useful for supporting wildlife-habitat mapping efforts.

Ecological systems are defined as “groups of plant community types that tend to co-occur within landscapes with similar ecological processes, substrates and/or environmental gradients” (Comer and others, 2003). As such, the TES definition emphasizes dominant vegetation type but also incorporates physical characteristics such as topographic position, hydrology, climate, and substrate. These characteristics also make TES units accessible for mapping using synoptic satellite imagery with ancillary data, including digital elevation models (DEMs) and derivatives (Lowry and others, 2006). Additional information about the TES Classification for the United States is available at <http://www.natureserve.org/publications/usEcologicalsystems.jsp> (accessed October 21, 2010).



**Figure 1.** Southwest Regional Gap Analysis Project (SWReGAP) map for the analysis area surrounding the Santa Cruz Watershed (modified from Lowry and others, 2006).

In addition to the existing SWReGAP map, detailed vegetation mapping for a portion of the Santa Cruz River riparian corridor was used to refine riparian vegetation units of interest (Sonoran Institute, 2008; Drake and others, 2009; Villarreal, 2009; Villarreal and others 2011b). Previous studies had recognized problems in the SWReGAP depiction of riparian vegetation types for this area, including overmapping Mesquite Bosque (Fonseca, 2008). Mesquite Bosque is a gallery forest found along the riparian flood plains of stream and river banks in the Sonoran Desert. The overmapping of this association is evident in figure 2, where Villarreal's detailed vegetation mapping (2009; fig. 2A) is compared with the National Land Cover Database (NLCD; fig. 2B; Homer and others, 2004; Fry and others, 2009), SWReGAP (fig. 2C; Lowry and others, 2006), and the new map of standardized ecosystems of the conterminous United States (fig. 2D; Sayre and others, 2008). The Santa Cruz River map of riparian vegetation and land cover developed by Villarreal (2009) for the National Park Service has a map accuracy of 83–95 percent for individual riparian vegetation classes.



**Figure 2.** Mapped riparian vegetation classes. A, 2004 Santa Cruz River map showing riparian vegetation and land cover (Villarreal, 2009). B, 2001 National Land Cover Dataset (NLCD) map; Homer and others, 2004; Fry and others, 2009. C, Southwest Regional Gap Analysis Project (SWReGAP) map. D, U.S. Geological Survey Standardized Terrestrial Ecosystems map.

The small segment of the Santa Cruz River riparian corridor shown in figure 2 contains a dense mix of Riparian Forest and Woodland and Mesquite Bosque. Most of these classes were unmapped in the standardized ecosystems map and significant portions were missed in the NLCD and the SWReGAP. The NLCD does not split out the Mesquite Bosque, and the SWReGAP maps all as Mesquite Bosque at the expense of Riparian Forest and Woodland. These distinctions between riparian vegetation types represent important differences in habitat for many wildlife species.

## Objective

The main objective of this study is to produce a detailed vegetation map of the Santa Cruz Watershed based on NatureServe's Terrestrial Ecological Systems (TES) mesoscale vegetation mapping units that seamlessly spans the international border for supporting ecosystem services evaluation and wildlife habitat modeling. An additional objective is to enhance the existing SWReGAP TES map, produced for the U.S. portion of the watershed (Lowry and others, 2006), by incorporating details of Santa Cruz River riparian corridor vegetation mapped by Villarreal (2009).

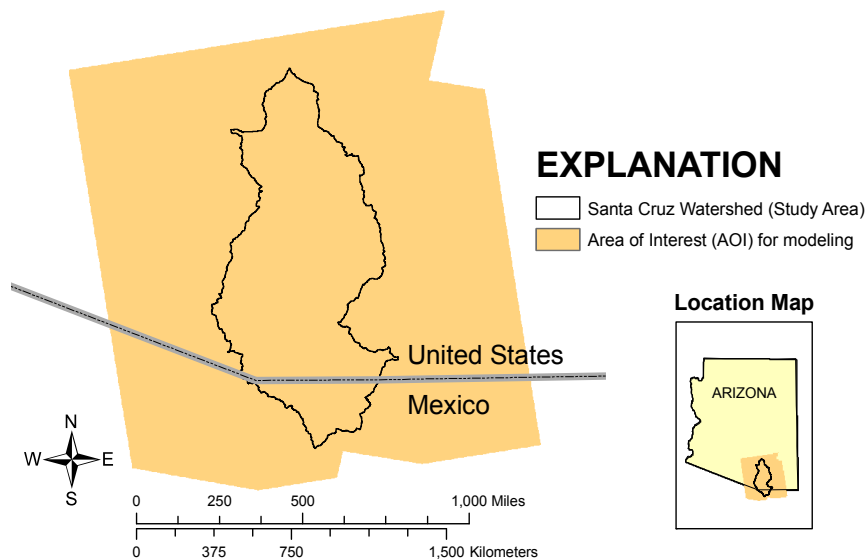


## Model Inputs

The CART modeling environment requires “ground truth” data used to train the classifier and independent data used as predictors. The training data are typically field data, for which the vegetation classes to be modeled are known to occur at specific locations on the ground. The independent data are typically environmental data layers that capture various characteristics of the landscape. One benefit of the CART modeling is that independent data can be either continuous, such as elevation or satellite reflectance, or categorical, such as landforms or soils.

The environmental datasets used for this project were Landsat-derived and DEM-derived, chosen on the basis of an inspection of existing SWReGAP modeling results and because they span the international border. Although other datasets, such as surficial geology, could be useful in the classification, a seamless dataset with suitable detail was not available for the binational area. Furthermore, the inclusion of Landsat Thematic Mapper (TM) data, with six reflectance bands ranging from visible to shortwave infrared, captures information about the soils and surficial geology (Jensen, 1996).

The study area is the Santa Cruz Watershed, but the modeling was performed within a larger area of interest (AOI) containing the watershed (fig. 3). This shape is the intersection of nine Landsat image footprints (spring, summer, and fall images) and a rectangular DEM subset from the seamless National Elevation Data (NED) 30-meter raster data available through the USGS at <http://gisdata.usgs.gov/website/seamless/>.



**Figure 3.** Study area. The area of interest (AOI) for modeling is the polygonal shape representing the intersection of the nine Landsat image footprints used in the model and a rectangular Digital Elevation Model (DEM) subset for the binational area.

## Independent Data

The Landsat images used by the SWReGAP modelers (<http://fws-nmcfwru.nmsu.edu/swregap/>, accessed October 21, 2010) were downloaded from the

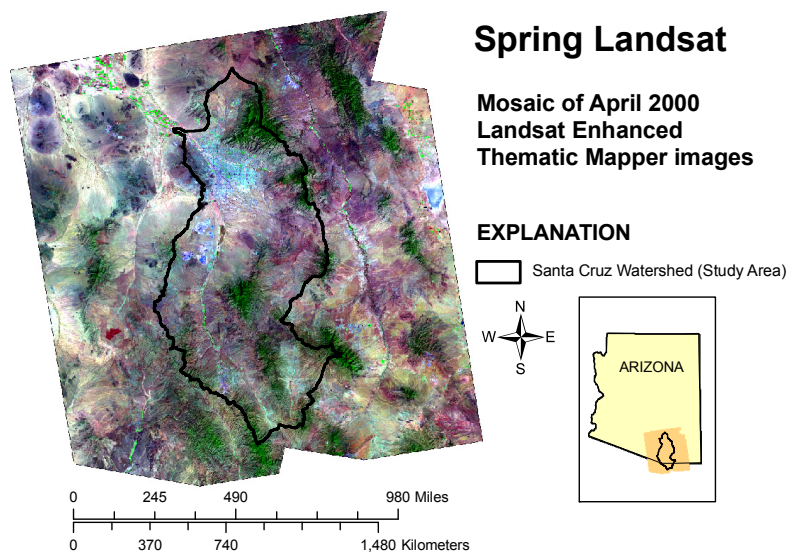
USGS Global Visualization Viewer (GLOVIS) Web site (<http://glovis.usgs.gov/>) and radiometrically corrected to top of atmosphere (TOA) reflectance using models created in Erdas Imagine 9.1 (Erdas, 2006). These images are from 1999 and 2000, coincident with the collection of the field data by the New Mexico GAP (NM-GAP) modeling team. After radiometric correction, the three contemporaneous Landsat images (P35R38, P36R37, and P36R38) for each of three seasons (spring, summer, and fall) were mosaiced in Erdas Imagine and clipped to the AOI.

A seasonal set of Landsat images is used to capture the dynamics of vegetation greenness throughout the year. These dynamics reflect the vegetation phenology, or seasonally recurring ecological events, such as vegetation green up, leafing or flowering, and senescence. The seasonal Landsat set captures information related to phenology and helps discriminate between vegetation communities on the basis of their distinctive dynamics. The Landsat images and derivatives used for modeling are as follows:

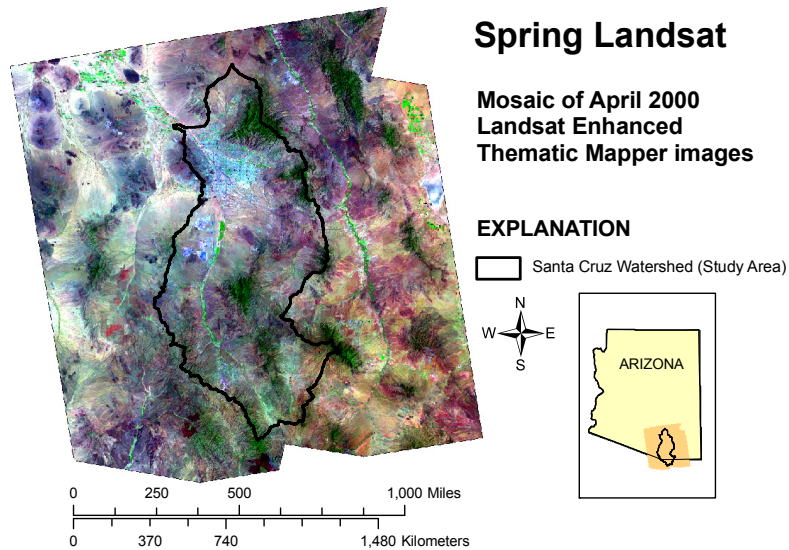
1. Landsat Mosaics. The seasonal mosaics are the original Landsat images, corrected to TOA, mosaiced using Erdas Imagine (specifying P36R38 as the reference image with no color balancing and no feathering), and clipped to the AOI.
  1. Spring Landsat Bands (six bands: blue (B), green (G), red (R), near-infrared (NIR), middle-infrared (MIR), and shortwave infrared (SWIR); fig. 4)
  2. Summer Landsat Bands (six bands: B,G,R,NIR,MIR,SWIR; fig. 5)
  3. Fall Landsat Bands (six bands: B,G,R,NIR,MIR,SWIR; fig. 6)
2. Tasseled Cap Indices. The Tasseled Cap algorithm transforms the original multiband TM data onto three orthogonal axes that represent overall “Brightness,” “Greenness,” and “Wetness” (Crist and Kauth, 1986). The seasonal tasseled cap images were calculated in Erdas Imagine (specifying the output stretched to 8-bit) from the mosaiced images and then clipped to the AOI.
  - a. Spring Tasseled Cap (three bands: Brightness, Greenness, and Wetness)
  - b. Summer Tasseled Cap (three bands: Brightness, Greenness, and Wetness)
  - c. Fall Tasseled Cap (three bands: Brightness, Greenness, and Wetness)
3. Normalized Difference Vegetation Index (NDVI) Images. The NDVI is an index derived from reflectance values of the R and NIR regions of the electromagnetic spectrum and is sensitive to various biophysical vegetation characteristics, such as biomass and percent cover (Huete and Jackson, 1987; Duncan and others, 1993). The formula is  $NDVI = (NIR - R) / (NIR + R)$ . NDVI values range from -1 to 1; non-land surfaces (such as water and snow) typically assume negative values and land surfaces typically assume positive values. As landscapes become more densely vegetated, the NDVI trends to 1. For this modeling, NDVI images were calculated in Erdas Imagine (specifying the output stretched to 8-bit) from the mosaiced images and then clipped to the AOI. Although NDVI images were not

used directly in the model, since the Tasseled Cap Greenness is an analogous metric, they were used to derive the texture images (number 4, this section).

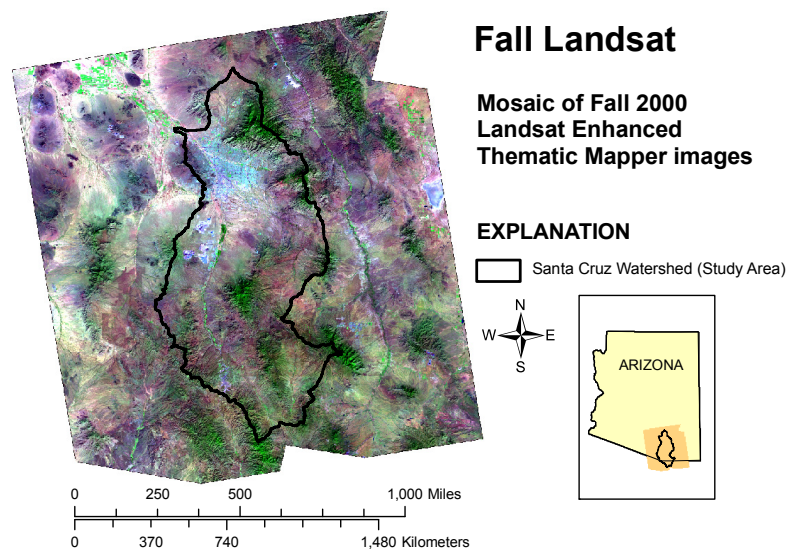
- a. Spring NDVI
  - b. Summer NDVI
  - c. Fall NDVI
4. Texture Images. These were calculated in Erdas Imagine (Interpreter, Spatial Enhancement, Texture, 3x3 Variance) from the mosaiced images and then clipped to the AOI.
    - a. April NDVI Texture: This derivative was included to capture information about the pattern of vegetation on the landscape.
    - b. October Red Band Texture: This derivative was included to capture information about the pattern of brightness on the landscape, especially to highlight urban/non-urban areas.



**Figure 4.** Spring Landsat. Mosaic of three Landsat Enhanced Thematic Mapper Plus (ETM+) images from spring 2000 (P35R38-April 13,2000; P36R37-April 12, 2000; and P36R38-April 12, 2000), displayed as: ETM band 7 (short-wave infrared) = Red, ETM band 4 (near-infrared) = Green, ETM band 1 (visible blue) = Blue. With this color combination, vegetation appears green, urban and barren areas appear blue, and sparsely-vegetated soils appear in shades of tan, pink and purple. The red patch visible to the west of the watershed is a recent fire scar.



**Figure 5.** Summer Landsat. Mosaic of three Landsat Enhanced Thematic Mapper Plus (ETM+) images from summer 2000 (P35R38-June 16, 2000; P36R37-June 15, 2000; and P36R38-June 15, 2000), displayed as: ETM band 7 (short-wave infrared) = Red, ETM band 4 (near-infrared) = Green, ETM band 1 (visible blue) = Blue. With this color combination, vegetation appears green, urban and barren areas appear blue, and sparsely-vegetated soils appear in shades of tan, pink and purple.

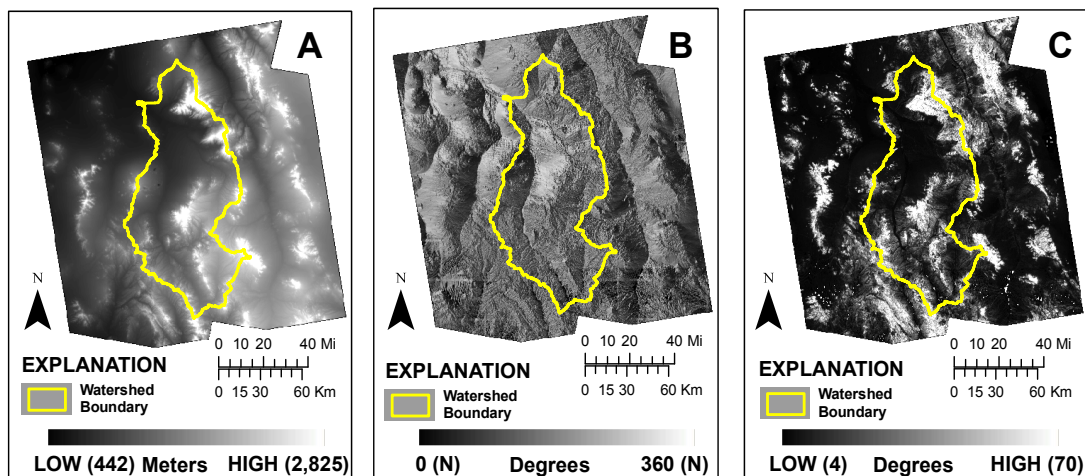


**Figure 6.** Fall Landsat. Mosaic of three Landsat Enhanced Thematic Mapper Plus (ETM+) images from fall 2000 (P36R37-October 19, 2000; P36R38-October 19, 2000; the October image P35R38 was too cloudy so the image used in the mosaic is an average of the two images: September 12, 2000 and November 13, 2000), displayed as: ETM band 7 (short-wave infrared) = Red, ETM band 4 (near-infrared) = Green, ETM band 1 (visible blue) = Blue. With this color combination, vegetation appears green, urban and barren areas appear blue, and sparsely-vegetated soils appear in shades of tan, pink and purple.

An additional suite of independent variables was created from topographic data for the region. A rectangular subset of the 30-meter resolution DEM containing the Santa Cruz Watershed study area was obtained through the USGS Earth Resources Observation and Science (EROS) Center (<http://edc.usgs.gov/>; fig. 7). The DEM-based raster data used for modeling and their derivations are as follows:

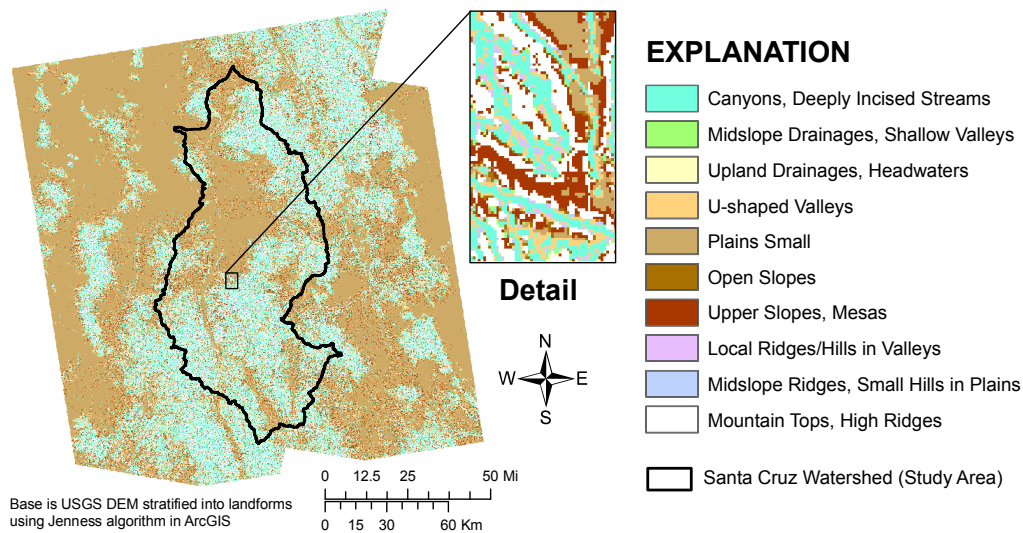
##### 5. DEM derivatives

- a. DEM: Original data, from the National Elevation Dataset (NED) 30-meter raster data available at: <http://gisdata.usgs.gov/website/seamless/> (fig. 7A).
- b. Aspect: Derived in Erdas Imagine (Interpreter, Topographic Analysis, Aspect) from the DEM and then clipped to the AOI (fig. 7B).
- c. Slope: Derived in Erdas Imagine (Interpreter, Topographic Analysis, Slope) from the DEM and then clipped to the AOI (fig. 7C).
- d. Landform: A stratification of the DEM into 10 landform classes as developed by Tagil and Jenness (2008; fig. 8). The algorithm that produced the Landform layer was accessed in the ArcGIS toolbox (Esri, 2010).
- e. Flow Accumulation: Derived using ArcGIS Spatial Analyst Hydrology tools. The DEM was first “Filled,” then “Flow Direction” was calculated, and finally “Flow Accumulation” was derived. Although the Flow Accumulation (fig. 9A) was not used directly in the modeling, it was used to derive the next layer.
- f. Flow Accumulation Difference of Focal Max and Mean: Calculated in Erdas Imagine using the Model Maker from the mosaiced images and then clipped to the AOI. Several variations on this type of neighborhood calculation were explored. The calculation that seemed to produce the desired result is subtracting the mean Flow Accumulation value within a 25x25 window from the maximum Flow Accumulation value within a 5x5 window. This difference was designed to pull out the areas within and near the topographic drainages to effectively constrain classification of riparian vegetation types. Figure 9B shows this derivative image.

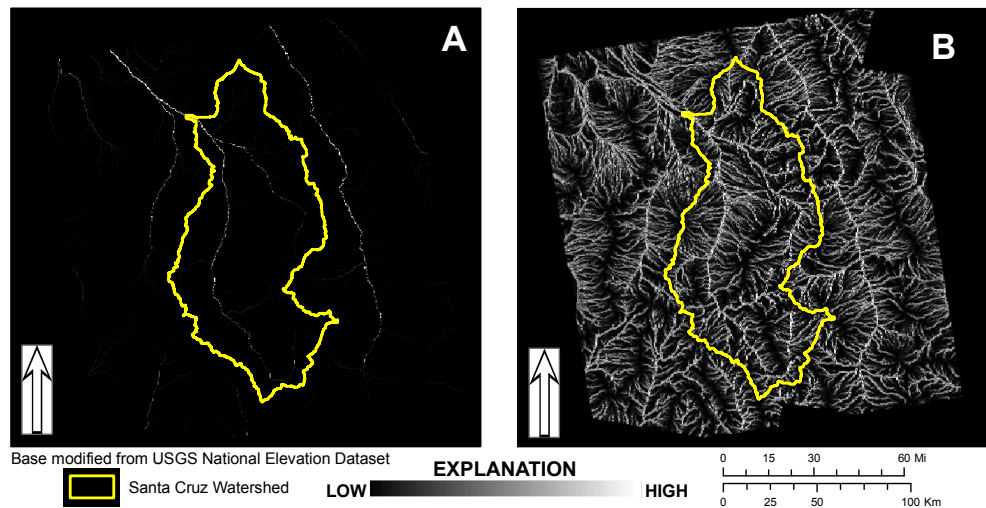


**Figure 7.** Digital Elevation Model (DEM) of the study area and its derivatives. *A*, DEM. *B*, Aspect. *C*, Slope. Elevation derivatives included in the model are Aspect (*B*) and Slope (*C*), both of which were calculated in Erdas Imagine (Erdas, 2006).





**Figure 8.** Landforms of the Santa Cruz Watershed.



**Figure 9.** Flow Accumulation (A) and derivatives (B).

### Training Data

Training data, which are the “ground truth” data used to train the classifier, are from three sources:

1. NM–GAP field data: SWReGAP training data for the New Mexico SWReGAP (NM–GAP) group’s ecoregion NM–3. These consist of polygons identified within their NM–3 mapping unit that intersect the AOI for this study.
2. Santa Cruz River riparian corridor mapping. The training data used for this map included field data collected between 2006 and 2008 (Villarreal, 2009), Quickbird multispectral satellite imagery, and historical aerial photographs.

3. SWReGAP classification. Initial CART modeling trials revealed that the available field data from SWReGAP field sites and Villarreal (2009) field sites were too limited within the AOI to produce a reasonable classification. Since our goal was to expand and refine the SWReGAP vegetation mapping to the full binational watershed, and we did not have the resources to collect additional field data, additional training data were extracted from the SWReGAP classification itself.

## **Methods**

### **Independent Data Preparation**

The environmental data layers described in the section “Model Inputs” were all clipped to the AOI and collected in a single folder for modeling. Although additional layers were derived, tested and considered, the layers shown are the ones used in the final model. An example of other layers considered is a Principal Components Analysis (PCA) of the original image bands, but this transformation is analogous to a Tasseled Cap transformation (they are both an orthogonalization of the bands into new uncorrelated bands). Furthermore, in contrast to PCA, the Tasseled Cap bands have an interpreted physical meaning of Brightness, Greenness, and Wetness.

### **Training Data Preparation**

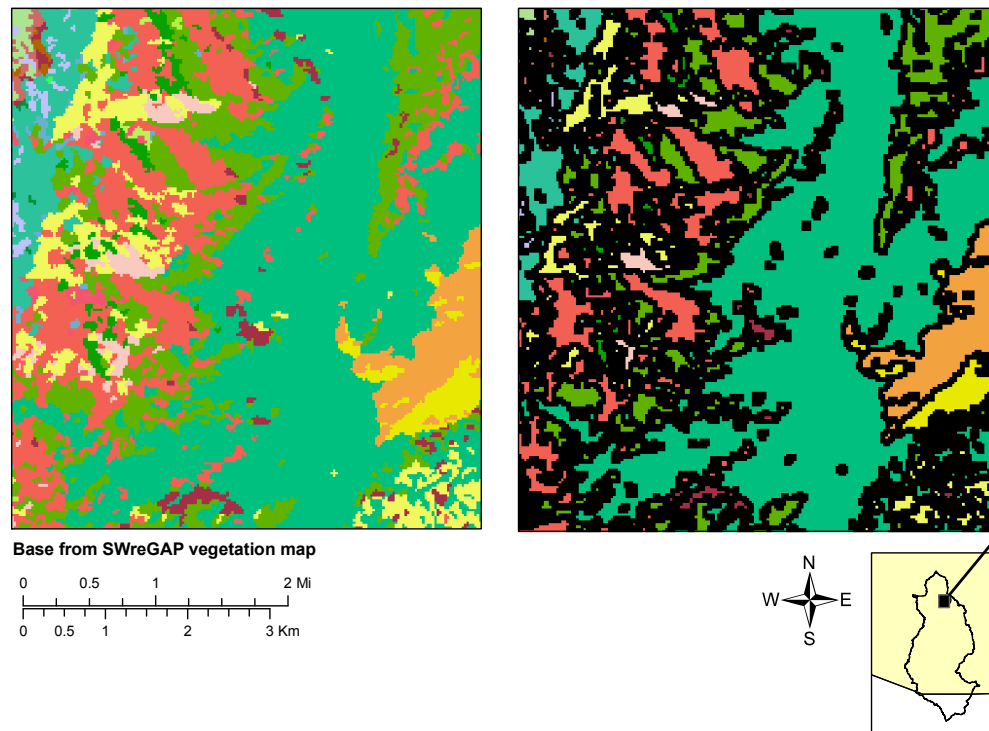
To simplify the classification process and enhance processing speed, we converted all training data into points. For the NM–GAP training sites, random points were gathered within the polygons by using a stratified-random sampling algorithm in Erdas Imagine specifying a minimum of 10 points per class. The points were then inspected on a 2006 Digital Orthophoto Quarter Quadrangle (DOQQ) and a few were deleted if they appeared to be unrepresentative of their labeled land-cover type (for example, the random point was located near a polygon edge). All of the riparian points were deleted because another data set (Villarreal, 2009) provided points within the Santa Cruz riparian corridor. A total of 447 training points were generated in this manner.

The training data Villarreal (2009) collected for 2006 (field data) and 1996 (interpreted aerial photos) were subset to those that were mapped identically for both time periods, to identify field data that existed as labeled in our intermediate time period of 1999–2000. The remaining training site polygon arcs were buffered by 60 meters, to exclude mixed-pixel edge areas, and random points were gathered within the reduced polygons by using a stratified-random sampling algorithm in Erdas Imagine, again specifying a minimum of 10 points per class. The points were then inspected on a 2006 DOQQ and a few were deleted if they appeared to be unrepresentative of their labeled land cover type. A total of 237 points were generated in this manner.

Mapped vegetation classes for the Santa Cruz River map were cross-walked to the SWReGAP TES units, requiring generalization of some classes. For example, “Riparian Woodland,” “Shrub Savanna,” “Shrubland,” and “Tree Savanna” were all merged into the SWReGAP “North American Warm Desert Riparian Woodland and Shrubland.” To

preserve ecological detail we considered important for wildlife habitat, we added a “Riparian Forest” class to the existing SWReGAP TES units.

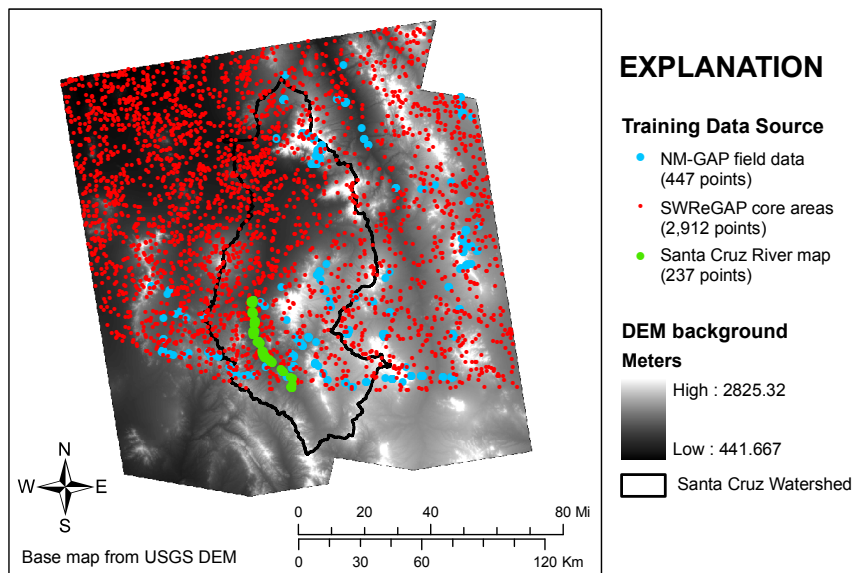
Additional training points were collected within the core areas of the SWReGAP polygons by using a stratified-random sampling algorithm in Erdas Imagine, again specifying a minimum of 10 points per class. The core areas of the polygons were extracted using the Erdas Imagine model maker by first calculating a Focal Diversity within a 3x3 window for the thematic image and then eliminating (setting to zero) all resulting values greater than 1 (fig. 10). A total of 2,912 points were generated in this manner.



**Figure 10.** Southwest Regional Gap Analysis Project (SWReGAP) core areas.

We removed all “Agriculture” points from the SWReGAP core data, for three reasons. First, SWReGAP did not use CART to model agriculture—it had been captured through heads-up digitizing or from other agency geographic information system (GIS) data. Second, the only agriculture evident in the Santa Cruz Watershed occurs along the riparian corridor and the SWReGAP includes this agriculture as well as upland center-pivot agriculture; these are distinctly different types of agriculture and would likely confound the CART classifier. Third, the Villarreal (2009) data include field points for the Santa Cruz riparian corridor agriculture, which are the data we preserved for training purposes. Figure 11 shows the locations and provenance of all the points used as training data.





**Figure 11.** Training points from New Mexico Gap Analysis Project (NM–GAP) field data, Villarreal (2009) field data, and random points from within the Southwest Regional Gap Analysis Project (SWReGAP) core areas.

### **Vegetation Classification**

The vegetation classification of the Santa Cruz Watershed was accomplished by using Classification And Regression Tree (CART) software (Quinlan, 1993, 1996). We accessed the CART tool developed by the USGS–NLCD for Erdas Imagine that is available free for download at <http://www.mrlc.gov/>. For each modeling run, we first used the “NLCD sampling tool” to create a “.data” file by inputting the independent data layers (all clipped to the AOI) and specifying the “training data” as a tab-delimited text file containing the three columns X, Y, and CODE (input to the tool as a .txt file without a header). The See5 program (installed as a stand-alone program outside of the Erdas toolbar) was then run, with the “.data” file specified as input and “construct classifier” executed to produce the output “.names” file. Several options are offered in the “construct classifier” menu. After some trial and error, we chose to “boost” 10 trials and chose Global Pruning (specifying Pruning CF 25 percent, minimum two cases) for the final classification runs. Once See5 is run, the NLCD tool “See5 classifier” is activated, choosing the See5 created “.names” file as input and specifying a “tree” classifier. The output is a classified image.

The final classification was the result of an iterative process. Early attempts, using only the field data from NM–GAP and Villarreal’s mapping, were unsuccessful and prompted the addition of SWReGAP core-area training sites. The next series of classifications produced an overmapping of riparian vegetation types outside of prominent drainage areas. This was controlled by creating the Flow Accumulation Difference of Focal Max and Mean (fig. 9B) for use as an additional independent data

layer and by digitizing additional training points from the SWReGAP classification in the areas incorrectly classified (234 points were added at this stage).

All TES classes were initially modeled, including anthropogenic and disturbed landscapes, such as Developed (urban) and Recently Mined or Quarried. Many of these nonvegetation classes tend to be quite heterogeneous, and although the CART environment produced reasonable results with the suite of independent layers we included, small patches of these types were identified by the CART classifier in improbable landscapes. Because of this, we followed the protocol of the original SWReGAP modelers, who captured these nonvegetation classes separately and applied them as an overlay on the CART-modeled vegetation classification. The SWReGAP modelers used a variety of methods, such as heads-up digitizing of Urban and Agriculture classes from high-resolution imagery and accessing other agency GIS data.

For our product, we accessed and overlaid Developed classes (high, medium, and low density) from the Land Cover product for 1999 produced under a parallel effort for this project (Villarreal and others, 2011a). In addition, the Recently Mined or Quarried class was accessed from the original SWReGAP product. Inspection of our model that included all TES classes revealed no obvious Recently Mined or Quarried landscapes in the Mexican portion of the watershed, so this overlay of the U.S. side only was considered reasonable.

The water class was produced as a combination of modeling and overlay. Training sites representing water were used in the CART classification to model a water class, but modeling resulted in clearly misclassified pixels in the interiors of some water bodies. To better delineate the water class, we accessed and overlaid the SWReGAP water class as an additional source of classified water pixels.

Further inspection of the vegetation classification revealed an abundance of mapped Chihuahuan Mixed Salt Desert Scrub in the foothills of the mountains as a transition between Apacherian-Chihuahuan Mesquite Upland Scrub and the Madrean Encinal. This did not seem reasonable on the basis of our field knowledge of the area and did not fit our interpretation of the vegetation as revealed in high resolution National Agricultural Imagery Program (NAIP) imagery. Only two field training sites of this vegetation type were collected in the watershed by the NM-GAP team (AZ071202ES13 and AZ071202ES13), and inspection of the photographs from the SWReGAP archives suggest these two sites do not represent a classic Chihuahuan Mixed Salt Desert Scrub assemblage. Given that this salt desert type is typical of saline basins in the Chihuahuan Desert and the CART model has placed it in the alluvial fans within our watershed, we consider it more accurate to classify these landscapes as an ecotone of Apacherian-Chihuahuan Mesquite Upland Scrub. We have, therefore, relabeled it as Apacherian-Chihuahuan Mesquite Upland Scrub in our final classification.

In summary, the CART tool was used to model only natural and vegetated landscapes. A 3x3 majority filter was applied to the output classification to eliminate

small isolated pixels that are generally considered noise. The Developed, Recently Mined or Quarried, and Water classes were then overlain, the modeled Chihuahuan Mesquite Upland Scrub class was relabeled Apacherian-Chihuahuan Mesquite Upland Scrub, and the final product was clipped to the Santa Cruz Watershed study area.

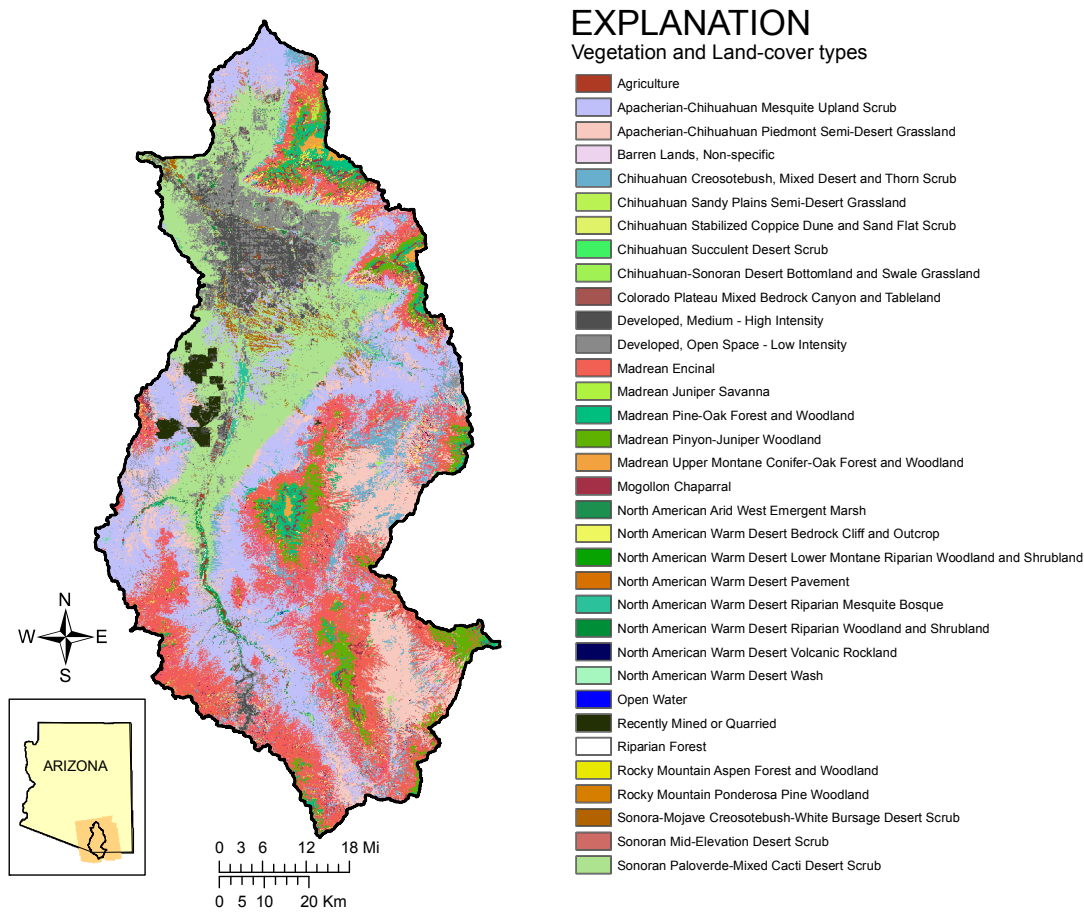
### **Accuracy Assessment**

Accuracy was quantified from a classification based on 80 percent of the training data and evaluated with the remaining 20 percent withheld as “testing” data. Following the protocol of the SWReGAP team, the final classification was created using 100 percent of the training data. A random sample to separate data into training and testing sets was generated for each data provenance (NM–GAP, Villarreal, SWReGAP core areas). A total of 641 points were withheld for testing: 89 from NM–GAP, 48 from Villarreal (2009), and 504 from SWReGAP core areas. After the elimination of nonvegetation training sites from the SWReGAP and NM–GAP sets and the addition of training sites in areas of misclassified riparian vegetation, a total of 3,448 training sites were used in the final classification.

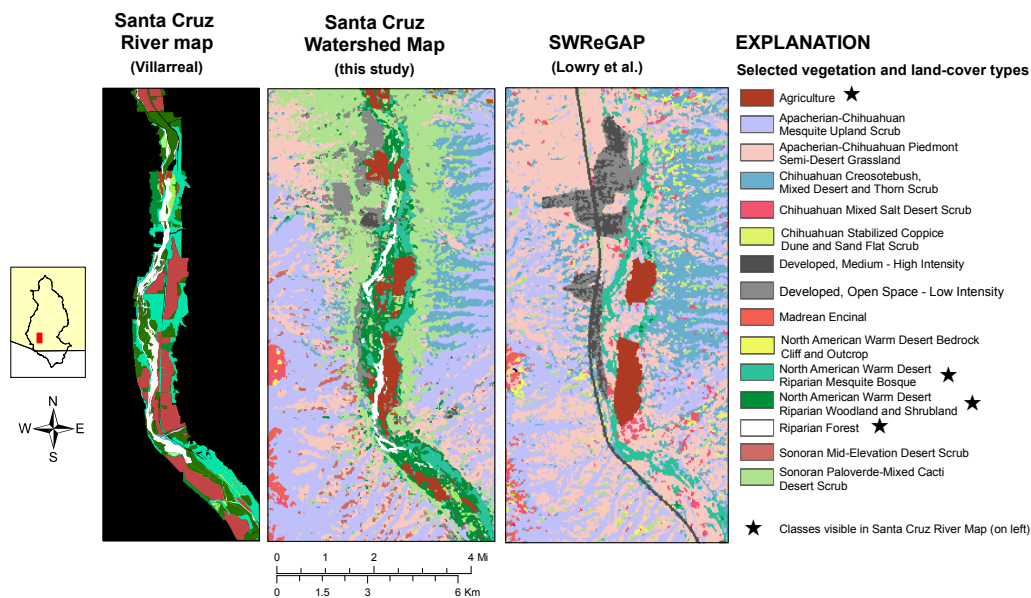
### **Results**

The final high-resolution binational map of the Santa Cruz River riparian habitat and watershed vegetation based on NatureServe TES units (Santa Cruz Watershed map) is shown in figure 12. The map is available for download as a georeferenced raster dataset in Erdas Imagine image format ([http://pubs.usgs.gov/of/2011/1143/of2011-1143\\_data.zip](http://pubs.usgs.gov/of/2011/1143/of2011-1143_data.zip)) with associated metadata ([http://pubs.usgs.gov/of/2011/1143/of2011-1143\\_metadata.txt](http://pubs.usgs.gov/of/2011/1143/of2011-1143_metadata.txt)). Visual comparison of the detailed Santa Cruz River map, the Santa Cruz Watershed map, and the SWReGAP map for a portion of the river near Tumacacori is shown in figure 13. The Santa Cruz Watershed map appears to be effective in representing units mapped by Villarreal (2009) in the Santa Cruz River map, and provides enhanced detail of the riparian vegetation units compared to the original SWReGAP product. The inclusion of the additional Riparian Forest class is especially striking and will provide useful data for wildlife habitat modelers. There are differences noted, such as the extent of agriculture, but some of these may be attributable the fact that the Santa Cruz River map represents conditions in 2006, whereas the data input to our model are from 1999 to 2000.

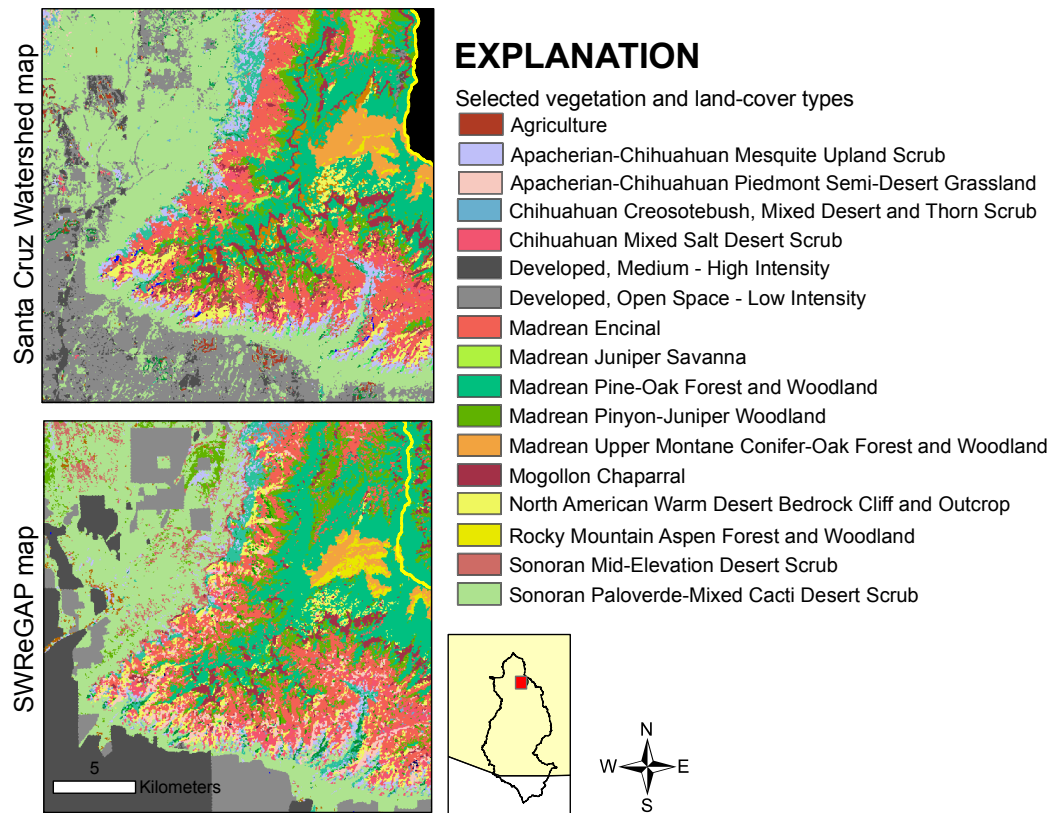
Figure 14 shows the Santa Cruz Watershed map and the SWReGAP for the Santa Catalina Mountains north of Tucson; the Santa Cruz Watershed map honors the overall vegetation patterns of the SWReGAP, which is reassuring given that the SWReGAP map was used as a major source of training data. For the riparian corridor area of Figure 13, the main difference between the Santa Cruz Watershed map and the SWReGAP map is the introduction of Sonoran Paloverde-Mixed Cacti Desert Scrub (light green), replacing some of the Apacherian-Chihuahuan Piedmont Semi-Desert Grassland (pale pink). Inspection of 1996 and 2005 aerial photographs shows shrub canopies are dominant in this area, supporting the classification as a desert scrub community (Santa Cruz Watershed map, this study) rather than a grassland community (SWReGAP).



**Figure 12.** Santa Cruz Watershed map of the riparian corridor and surrounding watershed.



**Figure 13.** Vegetation classifications in three maps of the riparian corridor.



**Figure 14.** Selected Terrestrial Ecological Systems (TES) units in the Santa Catalina Mountains shown in the Santa Cruz Watershed map and the Southwest Regional Gap Analysis Project (SWReGAP) map.

The overall accuracy of all vegetation classes modeled with the random 80-percent subset of the training points and evaluated with the withheld 20-percent test points was 78 percent (see tables 1 and 2). Because the final classification was created using all the training and testing points, the final map accuracy is expected to be better than the accuracy for this intermediate classification. Column totals of table 1 show producer's accuracies and row totals show user's accuracies. Units modeled with large amounts of training data have good validation results, for example Sonoran Paloverde-Mixed Cacti Desert Scrub has a user's accuracy of 97 percent. Poor accuracy results for the Sonoran Mid-Elevation Desert Scrub likely reflect both the fact that it is an ecologically generalized class as well as the lack and quality of training data. There were no field-based data for this class in our analysis area, so all training data were derived exclusively from the original SWReGAP classification.









**Table 1.** Error matrix for CART modeled vegetation classes.

[All classes with at least four testing points are shown. See table 2 for colors and descriptions of classes]

CLASS	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	Total	User	Commission	Kappa
1	3	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	6	50%	0.50	0.50
2	0	2	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	5	40%	0.60	0.40
3	1	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	75%	0.25	0.75
4	0	0	1	27	2	2	1	0	0	0	1	0	0	0	1	0	0	0	0	35	77%	0.23	0.76
5	0	0	0	4	9	1	0	0	0	0	0	0	0	0	0	0	0	0	0	14	64%	0.36	0.63
6	0	0	0	0	0	97	11	3	1	0	6	0	0	0	0	1	1	0	0	120	85%	0.15	0.81
7	0	0	0	0	0	4	33	8	0	0	5	0	0	0	0	0	0	0	0	50	66%	0.34	0.63
8	0	0	0	0	0	0	0	156	4	0	0	1	0	0	0	0	0	0	0	161	97%	0.03	0.96
9	0	0	0	0	0	0	0	8	28	0	0	0	0	0	0	0	1	0	0	37	76%	0.24	0.74
10	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	4	100%	0.00	1.00
11	0	0	0	2	0	9	6	4	0	0	78	2	0	0	0	0	0	0	0	101	77%	0.23	0.73
12	0	0	0	0	0	3	0	4	0	0	0	11	0	0	0	1	0	1	0	20	55%	0.45	0.54
13	0	0	0	0	0	0	0	2	0	0	0	0	8	1	0	0	0	0	0	11	73%	0.27	0.72
14	0	0	0	0	0	1	0	0	0	0	2	1	1	7	0	0	0	0	0	12	58%	0.42	0.58
15	0	0	2	1	3	3	1	0	0	0	0	0	0	0	8	0	0	0	0	18	44%	0.56	0.44
16	0	0	0	1	0	5	0	0	0	0	0	0	1	0	0	0	0	0	0	7	0%	1.00	0.00
17	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	1	0	0	4	25%	0.75	0.25
18	0	0	0	0	0	0	0	3	0	0	0	1	0	0	0	0	0	6	0	10	60%	0.40	0.59
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	3	4	75%	0.25	0.75
Total	4	2	6	36	15	126	52	188	36	4	95	16	10	8	9	2	3	8	3	623			

**Table 2.** Classification cross-walk

[Number on left is “Class” shown in table 1. Colors and descriptions are those used in figures 12, 13, and 14]

CLASS	Map Color	DESCRIPTION
1		North American Warm Desert Bedrock Cliff and Outcrop
2		North American Warm Desert Volcanic Rockland
3		Madrean Pine-Oak Forest and Woodland
4		Madrean Encinal
5		Mogollon Chaparral
6		Apacherian-Chihuahuan Mesquite Upland Scrub
7		Chihuahuan Creosotebush, Mixed Desert and Thorn Scrub
8		Sonoran Paloverde-Mixed Cacti Desert Scrub
9		Sonora-Mojave Creosotebush-White Bursage Desert Scrub
10		Sonora-Mojave Mixed Salt Desert Scrub
11		Apacherian-Chihuahuan Piedmont Semi-Desert Grassland and Steppe
12		North American Warm Desert Riparian Woodland and Shrubland
13		North American Warm Desert Riparian Mesquite Bosque
14		Chihuahuan-Sonoran Desert Bottomland and Swale Grassland
15		Madrean Pinyon-Juniper Woodland
16		Sonoran Mid-Elevation Desert Scrub
17		Barren Lands, Non-specific
18		Agriculture
19		Riparian Forest

Error matrix tables were also generated for only the field data testing points (not shown). Accuracy for only NM–GAP test points was 76 percent and only Villarreal test points was 60 percent. The combined accuracy for only the test points based on field data (NM–GAP and Villarreal) was 71 percent. Highest accuracy for Villarreal test points are for those classes that mapped into the classification scheme one-to-one, which are the North American Warm Desert Riparian Mesquite Bosque (100 percent) and Riparian Forest (75 percent) Lowest accuracies are for the North American Warm Desert Riparian Woodland and Shrubland (53 percent) and the Chihuahuan-Sonoran Desert Bottomland and Swale Grassland (50 percent), both of which were a many-to-one mapping, suggesting that the some of the Villarreal classes that were collapsed into single SWReGAP classes may not have been representative of the TES unit as defined by the NM–GAP mappers. The mapping accuracy may be improved by reevaluating the classification cross-walk and by eliminating Villarreal map units that are uncharacteristic of the NatureServe-defined TES units.



## Conclusions

Using a CART modeling environment, we produced a binational vegetation classification of the Santa Cruz River riparian habitat and watershed vegetation based on NatureServe TES units. The map is available for download as a georeferenced raster dataset in Erdas Imagine image format ([http://pubs.usgs.gov/of/2011/1143/of2011-1143\\_data.zip](http://pubs.usgs.gov/of/2011/1143/of2011-1143_data.zip)) with associated metadata ([http://pubs.usgs.gov/of/2011/1143/of2011-1143\\_metadata.txt](http://pubs.usgs.gov/of/2011/1143/of2011-1143_metadata.txt)). Environmental layers used as independent, predictor data were derived from a seasonal set of Landsat TM images (spring, summer, and fall) and from a 30-meter DEM grid. Because both sources of environmental data are seamless across the international border, they are particularly suited to this binational modeling effort. Training data were compiled from existing field data collected for a recent map produced by Villarreal (2009) for the Santa Cruz riparian corridor and data collected by the NM–GAP team for the original SWReGAP modeling effort. Additional training data were collected from core areas of the SWReGAP classification itself, allowing the extrapolation of the SWReGAP mapping into the Mexican portion of the watershed without collecting additional training data—an expensive and logistically challenging endeavor.

We assumed that the training data, which are exclusively from the U.S. portion of the watershed and region, are representative of the vegetation south of the border. For the relatively small study area of the Santa Cruz Watershed, this is a reasonable assumption. Although particular details of the vegetation mapping at a TES-level of resolution will benefit from inclusion of more field-verified data to train the model, this map serves as a reasonable representation of the vegetation patterns and communities within the Santa Cruz Watershed for the intended purpose of informing hydrologic modeling and wildlife habitat modeling within the Santa Cruz Watershed Ecosystem Portfolio Model (SCWEPM).

## References

- Comer, P., Faber-Langendoen, D., Evans, R., Gawler, S., Josse, C., Kittel, G., Menard, S., Pyne, M., Reid, M., Schulz, K., Snow, K., and Teague, J., 2003, Ecological systems of the United States—a working classification of U.S. terrestrial systems: NatureServe, Arlington, Va., available online at <http://www.natureserve.org/library/usEcologicalsystems.pdf>.
- Crist, E.P., and Kauth, R.J., 1986, The tasseled cap de-mystified: Photogrammetric Engineering and Remote Sensing, v. 52, no. 1, p. 81–86.
- Drake, S., Buckley, S., Villarreal, M.L., Studd, S., and Hubbard, A.J., 2009, Vegetation classification, distribution and mapping report—Tumacácori National Historical Park, Natural Resource Report NPS/SODN/NRR–2009/148: National Park Service, Fort Collins, Co.
- Duncan, J., Stow, D., Franklin, J., and Hope, A., 1993, Assessing the relationship between spectral vegetation indices and shrub cover in the Jornado Basin, New Mexico: International Journal of Remote Sensing, v. 14, no. 18, p. 3,395–3,416.
- Erdas, 2006, ERDAS IMAGINE 9.1, Leica Geosystems Geospatial Imaging, LLC, Georgia USA. August 26, 2006.
- ESRI, 2010, Topography tools for ArcGIS (9.3, 9.2, 9.1/9.0), accessed October 21, 2010, at <http://arcscrips.esri.com/details.asp?dbid=15996>.



- Fonseca, J., Jones, C., and Regan, J., 2008, Change in wetland cover, 1992–2001, Pima County, Arizona: 5th Annual Research Insights in Semiarid Ecosystems (RISE) Symposium, The University of Arizona, Tucson, October 11, 2008, poster presentation.
- Fry, J.A., Coan, M.J., Homer, C.G., Meyer, D.K., and Wickham, J.D., 2009, Completion of the National Land Cover Database (NLCD) 1992–2001 Land Cover Change Retrofit product: U.S. Geological Survey Open-File Report 2008–1379, 18 p., available online at <http://pubs.usgs.gov/of/2008/1379/>.
- Grossman, D.H., Faber-Langendoen, D., Weakley, A.S., Anderson, M., Bourgeron, P., Crawford, R., Goodin, K., Landaal, S., Metzler, K., Patterson, K.D., Pyne, M., Reid, M., and Sneddon, L., 1998, International classification of ecological communities—terrestrial vegetation of the United States: The Nature Conservancy, Arlington, Va.
- Hansen, M., Dubayah, R., and DeFries, R., 1996, Classification trees—an alternative to traditional land cover classifiers: *International Journal of Remote Sensing*, v. 17, p. 1,075–1,081.
- Homer, C.G., Huang, Chengquan, Yang, Limin, Wylie, B.K., and Coan, M.J., 2004, Development of a 2001 National Land Cover Database for the United States: *Photogrammetric Engineering and Remote Sensing*, v. 70, no. 7, p. 820–840.
- Huete, A.R., and Jackson, R.D., 1987, Suitability of spectral indices for evaluating vegetation characteristics on arid rangelands: *Remote Sensing of Environment*, v. 23, p. 213–232.
- Jensen, J.R., 1996, *Introductory digital image processing—a remote sensing perspective* (2d ed.): Upper Saddle River, N.J., Prentice-Hall, 316 p.
- Labiosa, W.B., Bernknopf, R., Hearn, P., Hogan, D., Strong, D., Pearlstine, L., Mathie, A.M., Wein, A.M., Gillen, K., and Wachter, S., 2009, The South Florida Ecosystem Portfolio Model—a map-based multicriteria ecological, economic, and community land-use planning tool: U.S. Geological Survey Scientific Investigations Report 2009–5181, 41 p.
- Lawrence, R.L., and Wright, A., 2001, Rule-based classification systems using Classification and Regression Tree (CART) analysis: *Photogrammetric Engineering & Remote Sensing*, v. 67, p. 1,137–1,142.
- Lowry, J.H., Jr., Ramsey, R.D., Thomas, K.A., Schrupp, D., Kepner, W., Sajwaj, T., Kirby, J., Waller, E., Schrader, S., Falzarano, S., Langs, L., Manis, G., Wallace, C., Schulz, K., Comer, P., Pohs, K., Rieth, W., Velasquez, C., Wolk, B., Boykin, K., O'Brien, L., Prior-Magee, J.S., Bradford, D., and Thompson, B., 2006, Land cover classification and mapping, chap. 2 of J.S. Prior-Magee and others, eds., *Southwest Regional Gap Analysis Final Report*: U.S. Geological Survey, Gap Analysis Program, Moscow, Idaho.
- Norman, L., Tallent-Halsell, N., Labiosa, W., Weber, M., McCoy, A., Hirschboeck, K., Callegary, J., van Riper III, C., and Gray, F., 2010, Developing an ecosystem services online decision support tool to assess the impacts of climate change and urban growth in the Santa Cruz Watershed—where we live, work, and play: *Sustainability*, v. 2, no. 7, p. 2,044–2,069, available online at <http://www.mdpi.com/2071-1050/2/7/2044/>.

- Quinlan, J.R., 1993, C4.5—programs for machine learning: San Mateo, Calif., Morgan Kaufmann, 302 p.
- Quinlan, J.R. 1996, Bagging, boosting and c4.5: Thirteenth National Conference on Artificial Intelligence, Portland, Oreg., 1996, Proceedings, v. 1, AAAI Press, p. 725–730.
- Sonoran Institute, 2008, Santa Cruz River Riparian Vegetation Mapping Project, Santa Cruz County, Arizona: Sonoran Institute, University of Arizona, 68 p., available online at [http://sonoraninstitute.org/library/recoreading/doc\\_download/724-santa-cruz-river-riparian-vegetation-map-final-report.html](http://sonoraninstitute.org/library/recoreading/doc_download/724-santa-cruz-river-riparian-vegetation-map-final-report.html).
- Sayre, R., Comer, P., Warner, H., and Cress, J., 2009, A new map of standardized terrestrial ecosystems of the conterminous United States: U.S. Geological Survey Professional Paper 1768, 17 p.
- Tagil, Sermin, and Jenness, Jeff, 2008, GIS-based automated landform classification and topographic, landcover and geologic attributes of landforms around the Yazoren Polje, Turkey: Journal of Applied Sciences, v. 8, p. 910–921.
- Villarreal, M.L., 2009, Land use and disturbance interactions in dynamic arid systems—multiscale remote sensing approaches for monitoring and analyzing riparian vegetation change: Tucson, Arizona, University of Arizona, Ph.D. dissertation, 167 p.
- Villarreal, M.L., Norman, L.M., Wallace, C.S.A., and van Riper, Charles, III, 2011, A multitemporal (1979-2009) land-use/land-cover dataset of the binational Santa Cruz Watershed: U.S. Geological Survey Open-File Report 2011-1131, 26 p. and database [<http://pubs.usgs.gov/of/2011/1131/>].
- Villarreal, M.L., Drake, S., Marsh, S.E and A.L. McCoy. 2011b. The influence of wastewater subsidy, flood disturbance and neighbouring land use on current and historical patterns of riparian vegetation in a semi-arid watershed. River Research and Applications, doi:10.1002/rra.1510.